IET Research Journals

IET Computer Vision SPECIAL ISSUE on Visual Domain Adaptation and Generalization

IET Journals
The Institution of Engineering and Technology

ISSN 1751-8644
doi: 0000000000
www.ietdl.org

ILGNet: Inception Modules with Connected Local and Global Features for Efficient Image Aesthetic Quality Classification using Domain Adaptation

Xin Jin[1], *Le Wu*[1], *Xiaodong Li*[1], *Xiaokun Zhang*[1], *Jingying Chi*[2], *Siwei Peng*[2], *Shiming Ge*[3]*, *Geng Zhao*[1], *Shuying Li*[4]

[1] *Department of Computer Science and Technology, Beijing Electronic Science and Technology Institute, Beijing 100070, China*
[2] *College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China*
[3] *Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China*
[4] *The 16th Institute, China Aerospace Science and Technology Corporation, Xi'an 710100, China*
* *E-mail: geshiming@iie.ac.cn*
`https://github.com/BestiVictory/ILGnet`

Abstract: In this paper, we address a challenging problem of aesthetic image classification, which is to label an input image as high or low aesthetic quality. We take both the local and global features of images into consideration. A novel deep convolutional neural network named ILGNet is proposed, which combines both the Inception modules and an connected layer of both Local and Global features. The ILGnet is based on GoogLeNet. Thus, it is easy to use a pre-trained GoogLeNet for large-scale image classification problem and fine tune our connected layers on an large scale database of aesthetic related images: AVA, i.e. *domain adaptation*. The experiments reveal that our model achieves the state of the arts in AVA database. Both the training and testing speeds of our model are higher than those of the original GoogLeNet.

1 Introduction

Shooting good photos needs years of practice for photographers. However, it is often easy for people to classify an image into high or low aesthetic quality. As shown in Fig. 1, the left image is often considered as with higher aesthetic quality than the right one.

Recently, smart phones, social networks and cloud computing boost the amount of images in the public or private cloud. People need a better way to manage their photos than ever before. A important ability of today's photo management software is to automatically recommend good photos from large amount of daily photos. Besides, aesthetic quality assessment can be used in the following scenarios:

1. When you search images in the Internet, the aesthetic assessment engine can help to give you the ones with high aesthetic quality;
2. Nowadays, one may use their smart phones to shoot many photos everyday. Then, they struggle to find good photos from hundreds of photos so as to share selected ones in their social network such as Facebook, We chat, etc. In this scenario, aesthetic quality assessment can help them to make initial selection;
3. New image beautification software could be inspired by aesthetic quality assessment;
4. Large-scale on-line E-commerce platform need automatic designing of logo, banner or production introduction. The aesthetic quality classification can help to delete the ones with low aesthetic quality;
5. Automatic typesetting magazines, presentation documents, and scientific papers can rely on the aesthetic quality classification engines;
6. Other domains such as architecture, graphics, industry design, fashion design can use aesthetic quality assessment to classify hundreds of works into low or high quality.

Today, image aesthetic quality classification is still a chanllenging problem. Typically, the following reasons make it challenging:

(a) (b)

Fig. 1: The left image (a) is often considered as with higher aesthetic quality than the right one (b).

- Two classes of high and low aesthetic qualities contain large intra class differences;
- Many high level aesthetic rules v.s. low level image features;
- The subjective nature of human rating on aesthetic qualities of images.

Thus, people from computer vision, computational photography and computational aesthetics make this topic hot. In their early work, they design hand-crafted aesthetic image features, which are fed into a classification model or a regression model. Generic image features are also used in aesthetic quality classification. Today, deep convolutional neural networks are designed specially for aesthetic quality classification.

Recently, deep learning technologies have boosted the performance of many computer vision tasks [1][2][3][4]. Google proposed the inception module used in deep neural network architecture [5]. The name of inception module are from the work of Lin et al [6]. The inception module can be considered as a logical culmination of [5]. It is inspired by Arora et al. [7] in theory. In ILSVRC 2014, the architecture with inception module shows its benefits. The performance was significantly raised in the classification and detection

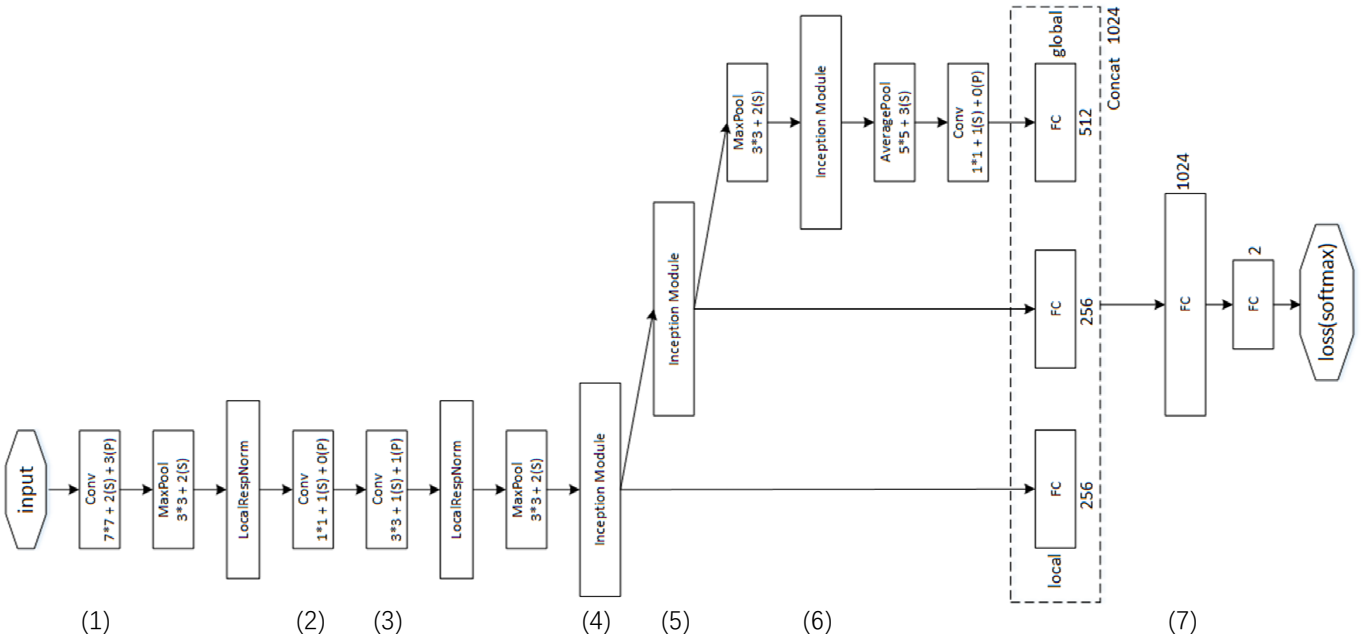

Fig. 2: The ILGNet architecture: Inception with connected Local and Global layers. We build this network on the first 1/3 part of GoogLeNetV1 [5] and batch normalization, which is a important feature of GoogLeNetV2 [11]. 1 pre-treatment layer and 3 inception modules are used. We use the first 2 inception modules to compute the local features and the last one to compute global features. Connecting intermediate layers directly to the output layers has show its value in recent work [8] [5]. Thus, we build a concat full connected layer of 1024 dimension which connect 2 layers of local features and a layer of global features. The output layer indicate the probability of low or high aesthetic quality. The ILGNet contains 13 layers with parameters and without counting pooling layers (4 layers). In Section 4, we use the labels (1)-(7) to demonstrate the visualization results.

challenges. However, in current literatures, inception modules has not been used in the aesthetic quality assessment to the best of our knowledge.

We propose to use inception modules for image aesthetics classification in this paper. A new deep convolutional neural network using Inception modules with connected Low and Global features is proposed, which is called ILGNet. Connecting intermediate layers directly to the output layers has show its value in recent work [8] [5]. In our ILGNet, the local features layers are connected to the global features layers. The ILGNet contains 13 layers with parameters and without counting pooling layers (4 layers). We use a pre-trained model on the ImageNet [9] as our initial model, which is trained for object classification of 1000 categories. Then, the inception modules are fixed and the connected local and global features layers are fine tuned on the AVA database, which is currently largest image aesthetics database [10]. We achieve the state of the art in the experiments on the AVA database [10]. Besides, the trained models and codes are available at github: `https://github.com/BestiVictory/ILGnet`.

The rest of this paper is organized as follows. In Section 2, we review the related work. In Section 3, we describe our proposed ILGNet in details. Then the experimental settings, results and comparisons with state-of-the-art methods are presented in Section 4. Finally, we give a conclusion in Section 5.

2 Previous Work

The related work of our task can be categorised into the traditional image quality assessment, the subjective image aesthetic quality assessment using hand-crafted features and deep learning.

2.1 Traditional Image Quality Assessment

Traditional image quality assessment is to assess the objective image quality, which may be distorted or influenced during the imaging, compression and transmission. Distortions such as ringing, blur, ghosting, smearing, blocking, mosaic, jerkiness are measured [12]. The human perception of aesthetics can not be well modeled by these low-level features and metrics.

2.2 Hand-crafted Features for Subjective Image Aesthetic Quality Assessment

Subjective image aesthetic quality assessment is to automatically distinguish an image to low or high aesthetic quality. Some of them can give a numerical assessment. They often contains the three steps in the following:

- A database of images is collected. Then they often manually label each with two label: *good* for images with high aesthetic quality, and *low* for images with low aesthetic quality. Some make psychological experiments so as to get numerical assessment for part of the images in the database.
- Image features for aesthetic quality assessment are designed such as simplicity, visual balance and rule of third [13][14][15] [16] [17] [18][19] [20] [21] [22][23] [24][25][26][27] [28] [29][30]. Generic image features which are previously used for object recognition are also used for aesthetic quality assessment, such as low level image features[31], bag of visual words [32][33], and Fisher Vector [34].
- Machine learning technologies such as random forest, support vector machine and boosting are used for image aesthetic quality assessment. They use the aesthetic database to train a classifier so as to classify an image into low or high aesthetic quality. They regress the human rating score to give a numerical assessment of the aesthetic quality of an image.

2.3 Subjective Image Aesthetic Quality Assessment using Deep Learning

Recently, deep learning technologies have boosted the performance of many computer vision tasks [1][2][3][4][35]. Deep belief network and deep convolutional neural network have been used for image aesthetics assessment. The performance has been significantly improved compared with traditional methods. [36][37] [38][39] [40] [41][12] [42] [43] [44] [45].

Most of the above work us the AlexNet architecture [46], which contains 8 layers with 5 convolutional layers and 3 full-connected layers or VGG [47]. Inspried by the good performance of GoogLeNet in the ImageNet, which argues that deeper architectures enable to capture large receptive field. We can extract local image features and the global features of the image layout. Connecting intermediate layers directly to the output layers has show its value in recent work [8] [5]. Both the local features and the global features can be extracted by inception modules. Thus, we change the GoogLeNet by connecting the intermediate local feature layers to the global feature layer.

3 ILGNet for Image Aesthetic Quality Classification

The details of the proposed ILGNet are described in this section. The ILGNet contains 13 layers with parameters and without counting pooling layers (4 layers). The network contains one pre-treatment layer and 3 inception modules. Two intermediate layers of local features are connected to a layer of global features, which makes a 1024 dimension concat layer. The output layer indicate the probability of low or high aesthetic quality. The basic ILGNet is built on the first 1/3 part of of GoogLeNetV1 [5] and batch normalization, which is a important feature of GoogLeNetV2 [11].

3.1 The Inception Module

The InceptionV1 module is proposed by GoogLeNetV1 [5]. The main ideas of the Inception module are:

1. Convolution kernels with different sizes represent receptive fields with difference sizes. This design means fusing features of different scales.
2. The kernel sizes are set to $1*1$, $3*3$ and $5*5$ so as to align the features conveniently. The stride is 1. The pad is set to 0, 1 ,2.
3. The features extracted by the higher layer are increasingly abstract. The receptive field involved by each feature is larger. Thus, the ratio of $3*3$ and $5*5$ kernels should be increased.

After InceptionV1, Google proposed InceptionV2 and InceptionV3, which adopt factorization of convolutions and improved normalization. Then, InceptionV4 considered the residue network, which surpassed its ancestor GoogLeNet on the ImageNet benchmark.

3.2 Image Aesthetic Quality Classification

The convolution layers inside ILGNet us rectified linear activation. The size of the input receptive field of ILGNet is 224×224 in color images with zero mean [5]. We use the first 2 inception modules to compute the local features and the last one to compute global features with 2 max pooling and 1 average pooling. Then, we build a concat full connected layer of 1024 dimension which connect 2 layers of local features (each layer is 256 dimension) and a layer of global features (512 dimension). The output layer is bypass a softmax layer to indicate the probability of low or high aesthetic quality.

The ILGnet is based on GoogLeNet. Thus, it is easy to use a pre-trained GoogLeNet for large-scale image classification problem and fine tune our connected layers on an large scale database of aesthetic related images: AVA [10], i.e. *domain adaptation*.

Table 1 The main training parameters of the Caffe package.

Parameters	AVA1 ($\delta = 0$)	AVA1 ($\delta = 1$)	AVA2
base_ lr	0.0001	0.00001	0.00001
lr_ policy	"step"	"step"	"step"
stepsize	100000	19000	13325
gamma	0.96	0.96	0.96
max_ iter	475000	760000	533000
momentum	0.9	0.9	0.9
weight_ decay	0.0002	0.0002	0.0002

4 Experimental Results

We test the effectiveness of our ILGNet in the public AVA datebase [10], which is specially designed for aesthetics analysis. The comparison experiments with the state of the art methods on aesthetic quality classification are shown in this section. Most of them use deep convolutional neural networks. The main training parameters of the Caffe package [48] are listed in Table 1.

4.1 Database and Comparison Protocols

The Aesthetic Visual Analysis database [10] is a list of image ids from DPChallenge.com, which is a on-line photography social network. There are total 255,529 photos, each of which is rated by 78-549 persons, with an average of 210. The range of the scores rated by human is 1-10. We use the same protocols to those of previous work. They often use two sub database of AVA.

- AVA1: The score of 5 are chosen as the threshold to distinguish the AVA to high (good) and low (bad) aesthetics quality. 74,673 images are labelled as bad photos. 180,856 are labelled as good photos. We randomly split the AVA database into training set (234,599) and testing set (19,930) [10][41] [44][42][39][38][12].
- AVA2: The images in the AVA database are sorted according to their mean scores of the aesthetic quality. Then the top 10% images are labelled as good. The bottom 10% are labelled as bad. Thus, there are totally 51,106 images from AVA database. The 51,106 images are randomly divided into 2 sets with equal numbers, which are the training set and testing set respectively [15][49][13][14][34][40][50][41].

4.2 Classification Results

As shown in Fig. 3, We use the trained ILGNet to label images with good or bad, which indicates high or low aesthetic quality, respectively. Differences between low-aesthetic images and high-aesthetic images heavily lie in the amount of textures and complexity of the entire image [39].

The original ILGNet is build on the first 1/3 of GoogLeNet V1, as shown in Fig. 2. We add batch normalization (GoogLeNet V2 [5] features), which form our ILGNet-Inc.V1-BN. After that we further build our ILGNet on the first 1/3 of recent GoogLeNet V3 [53] and V4 [54], which form our ILGNet-Inc.V3 and ILGNet-Inc.V4. The test results in the AVA1 database are shown in Table 2. Our ILGNet-Inc.V4 outperforms the other DCNN based methods and achieve the state of the art accuracy: 82.66%.

The above is the case of $\delta = 1$. Similar results are shown when $\delta = 1$. In the original test protocol [10], they set $\delta = 1$ in the training set, there are 7,500 low-quality images and 45,000 high-quality images. For the testing images, they fix δ to 0, regardless what δ is used for training. We have tested five network architectures on $\delta = 0$ and $\delta = 1$. The results are shown in Table 2. The ambiguity image samples are removed from the training set. Ambiguity images are still in the test set. Thus, the decreasing of accuracy is reasonable. We still achieve the state of the arts performance when delta =1.

Good

Bad

Fig. 3: We use the trained ILGNet to label images with good or bad, which indicates high or low aesthetic quality, respectively.

Table 2 The Classification Accuracy in AVA1 database.

Methods	$\delta = 0$	$\delta = 1$
Traditional method[10]	66.70%	67.00%
RAPID [37]	69.91%	71.26%
RAPID-E [39]	74.46%	73.70%
Multi-patch [38]	75.41%	–
AROD [51]	75.83%	–
Multi-scene [41]	76.94%	–
Comp.-prev. [12]	77.10%	76.10%
AADB [42]	77.33%	–
BDN [44]	78.08%	77.27%
Semantic-based [43]	79.08%	76.04%
A-Lamp [45]	82.5%	–
ILGNet-without-Inc.	**75.29%**	**73.25%**
1/3 GoogLeNetV1-BN	**80.74%**	**79.09%**
ILGNet-Inc.V1-BN	**81.68%**	**80.71%**
ILGNet-Inc.V3	**81.71%**	**80.65%**
ILGNet-Inc.V4	**82.66%**	**80.83%**

Table 3 The Classification Accuracy in AVA2 database.

Methods	Accuracy
Subject-based [15]	61.49%
EfficientAssess [49]	68.13%
Generic-based [34]	68.55%
Compt.-based [13]	68.67%
High-level [14]	71.06%
Multi-level [40]	78.92%
Query-dependent [52]	80.38%
DCNN-Aesth-SP [50]	83.52%
Multi-scene [41]	84.88%
ILGNet-without-Inc.	**79.64%**
1/3 GoogLeNetV1-BN	**82.26%**
ILGNet-Inc.V1-BN	**85.50%**
ILGNet-Inc.V3	**85.51%**
ILGNet-Inc.V4	**85.53%**

Table 4 The Efficiency Comparison in AVA1 database.

Methods	Accuracy $\delta = 0$	Training Time	Test Time
Full GoogLeNetV1-BN	**82.36%**	16 days	0.84s
2/3 GoogLeNetV1-BN	81.72%	11 days	0.57s
1/3 GoogLeNetV1-BN	80.74%	**4 days**	0.33s
ILGNet-Inc.V1-BN	81.68%	**4 days**	**0.31s**

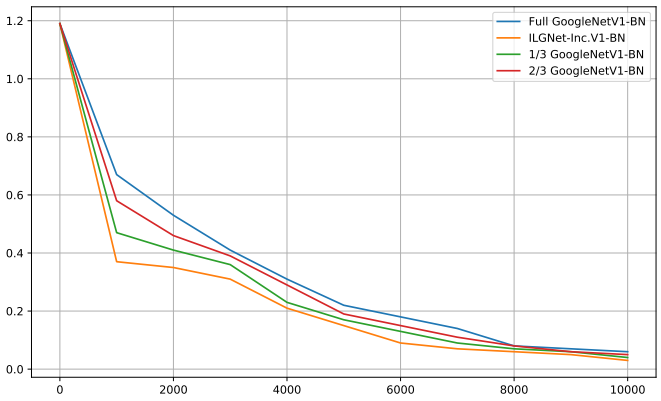

Fig. 4: Loss vs. epoch of our ILGNet-Inc.V1-BN, 1/3, 2/3 and full GoogLeNetV1-BN. in AVA1 database.

To verify the effectiveness of inception module, we test a modified network of ILGNet-Inc.V1-BN: the ILGNet-without-Inc., in which we replace all the inception module with corresponding ordinal convolutional layer that is adaptive with the original pre and next layers. The performance (75.29%) of this INGNet-without-Inc. is significantly worse than that (81.68%) of the ILGNet-Inc.V1-BN. This verifies the usefulness of the inception module in capture features of both local patch and global view.

To verify the effectiveness of the connected local and global layer, we compare our ILGNet-Inc.V1-BN with the first 1/3 of original GoogLeNet with batch normalization: 1/3 GoogleNetV1-BN. The performance (80.74%) of the 1/3 GoogleNetV1-BN on AVA1 is also worse than that (81.68%) of the ILGNet-Inc.V1-BN. This verifies the usefulness of our proposed connected local and global layer.

The test results in the AVA2 database are shown in Table 3. Our ILGNet-Inc.V4 outperforms the other DCNN based methods and achieve the state of the art accuracy: 85.53%.

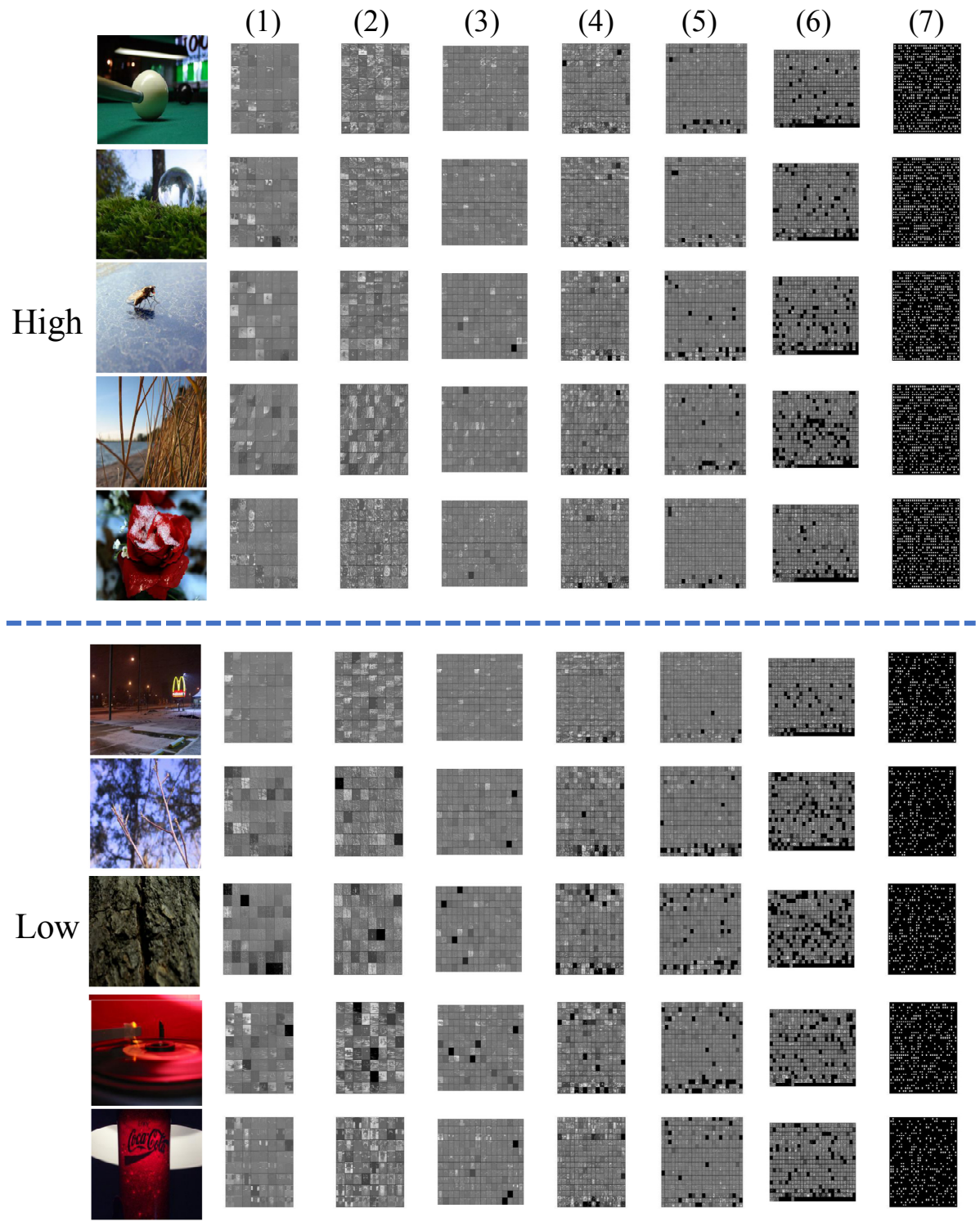

Fig. 5: The extracted features using the ILGNet-Inc.V1-BN of good and bad photos. The labels of (1)-(7) means the same in Fig. 2. We have an interesting observation that in the last layer, the density of the active features are often higher in the ones with high aesthetic quality than those with low aesthetic quality.

4.3 The Efficiency Comparison

We take the ILGNet-Inc.V1-BN as an example to compare the efficiency with the first 1/3, 2/3 and full GoogLeNetV1 plus batch normalization. The time costs are summarized in Table 4. The test time is the average time on the test set of AVA1 and a Nvidia GTX980ti card. The time cost of both training and test of the ILGNet-Inc.V1-BN are significantly less than those of full GoogLeNetV1-BN with only a little reduction of the classification accuracy. This makes the aesthetic assessment model more easily to be integrated into mobile and embedded systems.

The performance of our ILGNet-Inc.V1-BN is better than that of 1/3 GoogLeNetV1-BN. The training and test times of our ILGNet-Inc.V1-BN is similar as those of 1/3 GoogLeNetV1-BN. This is because that our ILGNet-Inc.V1-BN is built on the 1/3 GoogLeNetV1-BN, which has similar computational efficiency as ours. With our strategy of connected local and global layer, our ILGNet-Inc.V1-BN can even achieve nearly the same performance (81.68%) to that (81.72%) of 2/3 GoogleNetV1-BN. While the training and test times of 2/3 GoogleNetV1-BN are much more than those of our ILGNet-Inc.V1-BN. In addition, we show the loss vs. epoch curves in Fig. 4. Our ILGNet-Inc.V1-BN achieve the fastest convergence speed, which further verifies the efficiency of our method.

4.4 The Features Visualization

The extracted features using the ILGNet-Inc.V1-BN are visualized in Fig. 5 for images with high and low aesthetic quality. The proposed ILGNet-Inc.V1-BN can be used to compute the low level features and high level features. The connected layer of local and global features are shown at last. It can be observed that the last feature maps are nearly binary pattens. We have an interesting observation that in the last layer, the density of the active features are often higher in the ones with high aesthetic quality than those with low aesthetic quality. This verifies that the extracted features can well represent the aesthetic quality.

5 Conclusion and Discussion

We propose a new DCNN called ILGNet for subjective image aesthetic quality classification. The ILGNet is derived from part of GoogLeNet. Thus, it can be used for domain adaptation from image classification to image aesthetic quality classification. The bottom features are shared for this two tasks. The high level features together with 2 inception modules are fine tuned for aesthetic quality classification. We fixed the shared inception layers of a pre-trained GoogLeNet model on the ImageNet [9] and fine tune the connected layer on the AVA database [10]. The proposed ILGNet outperforms the state of the art methods in AVA database.

In the future work, we will address the following problems.

Hyperparameter. We hope that some architecture parameters such as the number of layers and the number of nodes on the full connected layers can also be automatically determined from the training on large-scale aesthetic database.

Composition. Now the input image is scaled to a fixed size of 224*224, which loses high quality local image patches and destroys the composition aesthetics of the original image. In the future work, we will use technologies such as spatial pyramid pooling to handle this limitation.

Database Bias. Because of the bias of AVA database (the number of high quality images is higher than that of low quality images), we will explore other aesthetic criteria such as numerical assessment or ranking in the future.

Consensus. The aesthetic assessment is a subjective task in nature. The mean score of an image can describe the overall impression to some extent with consensus. In the future work, we need to assess the aesthetic quality from more views such as score distribution, photography attributes, aesthetic caption.

Acknowledgments

This work is partially supported by the National Natural Science Foundation of China (Grant Nos. 61402021, 61401228, 61402463, 61772513), the Science and Technology Project of the State Archives Administrator (Grant No. 2015-B-10), the open funding project of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (Grant No. BUAA-VR-16KF-09), the Fundamental Research Funds for the Central Universities (Grant No. 3122014C017), the China Postdoctoral Science Foundation (Grant No. 2015M581841), and the Postdoctoral Science Foundation of Jiangsu Province (Grant No. 1501019A).

Parts of this paper have previously appeared in our previous work [55]. This is the extended journal version of that conference paper. The main differences between this archival version and the conference version are:

1. The title has been changed to capture essential idea of our work.
2. A clearer explanation of the power of the Inception module.
3. The performance of our ILGNet is increased to the state of the art (79.25% to 82.66% with the new network combination ILGNet-Inc.V4 reported in this version).
4. More experimental results are shown and more related methods (including the state of the art method published after the publication of our conference paper) are compared.
5. We have published our newly trained models and codes at `https://github.com/BestiVictory/ILGnet`

6 References

1 Q. Wang, J. Gao, and Y. Yuan, "A joint convolutional neural networks and context transfer for street scenes labeling," *IEEE Transactions on Intelligent Transportation Systems*, vol. PP, no. 99, pp. 1–14, 2017.

2 Q. Wang, J. Gao, and Y. Yuan, "Embedding structured contour and location prior in siamesed fully convolutional networks for road detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. PP, no. 99, pp. 1–12, 2017.

3 Jia Li, Changqun Xia, and Xiaowu Chen, "A benchmark dataset and saliency-guided stacked autoencoders for video-based salient object detection," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 349–364, 2018.

4 Xin Jin, Le Wu, Xiaodong Li, Siyu Chen, Siwei Peng, Jingying Chi, Shiming Ge, Chenggen Song, and Geng Zhao, "Predicting aesthetic score distribution through cumulative jensen-shannon divergence," *ArXiv, AAAI 2018*, vol. abs/1708.07089, 2017.

5 Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR, Boston, MA, USA, June 7-12, 2015*, 2015, pp. 1–9.

6 Min Lin, Qiang Chen, and Shuicheng Yan, "Network in network," *CoRR*, vol. abs/1312.4400, 2013.

7 Sanjeev Arora, Aditya Bhaskara, Rong Ge, and Tengyu Ma, "Provable bounds for learning some deep representations," in *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*, 2014, pp. 584–592.

8 Michael Maire, Stella X. Yu, and Pietro Perona, "Reconstructive sparse code transfer for contour detection and semantic labeling," in *Computer Vision - ACCV - 12th Asian Conference on Computer Vision, Singapore, Singapore, November 1-5, 2014, Revised Selected Papers, Part IV*, 2014, pp. 273–287.

9 Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Fei-Fei Li, "Imagenet: A large-scale hierarchical image database," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 20-25 June 2009, Miami, Florida, USA*, 2009, pp. 248–255.

10 Naila Murray, Luca Marchesotti, and Florent Perronnin, "AVA: A large-scale database for aesthetic visual analysis," in *IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, June 16-21, 2012*, 2012, pp. 2408–2415.

11 Sergey Ioffe and Christian Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, 2015, pp. 448–456.

12 Long Mai, Hailin Jin, and Feng Liu, "Composition-preserving deep photo aesthetics assessment," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

13 Ritendra Datta, Dhiraj Joshi, Jia Li, and James Ze Wang, "Studying aesthetics in photographic images using a computational approach," in *Computer Vision - ECCV, 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006, Proceedings, Part III*, 2006, pp. 288–301.

14 Yan Ke, Xiaoou Tang, and Feng Jing, "The design of high-level features for photo quality assessment," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 17-22 June 2006, New York, NY, USA*, 2006, pp. 419–426.

15 Yiwen Luo and Xiaoou Tang, "Photo and video quality evaluation: Focusing on the subject," in *Computer Vision - ECCV, 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part III*, 2008, pp. 386–399.

16 Congcong Li and Tsuhan Chen, "Aesthetic visual quality assessment of paintings," *J. Sel. Topics Signal Processing*, vol. 3, no. 2, pp. 236–252, 2009.

17 Subhabrata Bhattacharya, Rahul Sukthankar, and Mubarak Shah, "A framework for photo-quality assessment and enhancement based on visual aesthetics," in *Proceedings of the 18th International Conference on Multimedia 2010, Firenze, Italy, October 25-29, 2010*, 2010, pp. 271–280.

18 Wei Jiang, Alexander C. Loui, and Cathleen Daniels Cerosaletti, "Automatic aesthetic value assessment in photographic images," in *IEEE International Conference on Multimedia and Expo, ICME 2010, 19-23 July 2010, Singapore*, 2010, pp. 920–925.

19 Congcong Li, Andrew C. Gallagher, Alexander C. Loui, and Tsuhan Chen, "Aesthetic quality assessment of consumer photos with faces," in *Proceedings of the International Conference on Image Processing, ICIP 2010, September 26-29, Hong Kong, China*, 2010, pp. 3221–3224.

20 Xin Jin, Mingtian Zhao, Xiaowu Chen, Qinping Zhao, and Song Chun Zhu, "Learning artistic lighting template from portrait photographs," in *Computer Vision - ECCV 2010, 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part IV*, 2010, pp. 101–114.

21 Douglas Gray, Kai Yu, Wei Xu, and Yihong Gong, "Predicting facial beauty without landmarks," in *Computer Vision - ECCV 2010 - 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part VI*, 2010, pp. 434–447.

22 Xiaowu Chen, Xin Jin, Hongyu Wu, and Qinping Zhao, "Learning templates for artistic portrait lighting analysis," *IEEE Trans. Image Processing*, vol. 24, no. 2, pp. 608–618, 2015.

23 Sagnik Dhar, Vicente Ordonez, and Tamara L. Berg, "High level describable attributes for predicting aesthetics and interestingness," in *The 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, Colorado Springs, CO, USA, 20-25 June 2011*, 2011, pp. 1657–1664.

24 Dhiraj Joshi, Ritendra Datta, Elena A. Fedorovskaya, Quang-Tuan Luong, James Ze Wang, Jia Li, and Jiebo Luo, "Aesthetics and emotions in images," *IEEE Signal Process. Mag.*, vol. 28, no. 5, pp. 94–115, 2011.

25 Masashi Nishiyama, Takahiro Okabe, Imari Sato, and Yoichi Sato, "Aesthetic quality classification of photographs based on color harmony," in *The 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, Colorado Springs, CO, USA, 20-25 June 2011*, 2011, pp. 33–40.

26 Wei Luo, Xiaogang Wang, and Xiaoou Tang, "Content-based photo quality assessment," in *IEEE International Conference on Computer Vision, ICCV 2011, Barcelona, Spain, November 6-13, 2011*, 2011, pp. 2206–2213.

27 Xiaoou Tang, Wei Luo, and Xiaogang Wang, "Content-based photo quality assessment," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 1930–1943, 2013.

28 Ou Wu, Weiming Hu, and Jun Gao, "Learning to predict the perceived visual quality of photos," in *IEEE International Conference on Computer Vision, ICCV 2011, Barcelona, Spain, November 6-13, 2011*, 2011, pp. 225–232.

29 Shehroz S. Khan and Daniel Vogel, "Evaluating visual aesthetics in photographic portraiture," in *Computational Aesthetics 2012: Eurographics Workshop on Computational Aesthetics, Annecy, France, 4-6 June 2012. Proceedings*, 2012, pp. 55–62.

30 Yuzhen Niu and Feng Liu, "What makes a professional video? A computational aesthetics approach," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 22, no. 7, pp. 1037–1049, 2012.

31 Hanghang Tong, Mingjing Li, HongJiang Zhang, Jingrui He, and Changshui Zhang, "Classification of digital photos taken by photographers or home users," in *PCM, 5th Pacific Rim Conference on Multimedia, Tokyo, Japan, November 30 - December 3, 2004, Proceedings, Part I*, 2004, pp. 198–205.

32 Hsiao-Hang Su, Tse-Wei Chen, Chieh-Chi Kao, Winston H. Hsu, and Shao-Yi Chien, "Scenic photo quality assessment with bag of aesthetics-preserving features," in *Proceedings of the 19th International Conference on Multimedia 2011, Scottsdale, AZ, USA, November 28 - December 1, 2011*, 2011, pp. 1213–1216.

33 Hsiao-Hang Su, Tse-Wei Chen, Chieh-Chi Kao, Winston H. Hsu, and Shao-Yi Chien, "Preference-aware view recommendation system for scenic photos based on bag-of-aesthetics-preserving features," *IEEE Trans. Multimedia*, vol. 14, no. 3-2, pp. 833–843, 2012.

34 Luca Marchesotti, Florent Perronnin, Diane Larlus, and Gabriela Csurka, "Assessing the aesthetic quality of photographs using generic image descriptors," in *IEEE International Conference on Computer Vision, ICCV 2011, Barcelona, Spain, November 6-13, 2011*, 2011, pp. 1784–1791.

35 Jianzhe Lin, Chen He, Z. Jane Wang, and Shuiying Li, "Structure preserving transfer learning for unsupervised hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, pp. 1656–1660, 2017.

36 Sergey Karayev, Matthew Trentacoste, Helen Han, Aseem Agarwala, Trevor Darrell, Aaron Hertzmann, and Holger Winnemoeller, "Recognizing image style," in *British Machine Vision Conference, BMVC 2014, Nottingham, UK, September 1-5, 2014*, 2014.

37 Xin Lu, Zhe Lin, Hailin Jin, Jianchao Yang, and James Zijun Wang, "RAPID: rating pictorial aesthetics using deep learning," in *Proceedings of the ACM International Conference on Multimedia, MM'14, Orlando, FL, USA, November 03 - 07, 2014*, 2014, pp. 457–466.

38 Xin Lu, Zhe Lin, Xiaohui Shen, Radomír Mech, and James Zijun Wang, "Deep multi-patch aggregation network for image style, aesthetics, and quality estimation," in *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015*, 2015, pp. 990–998.

39 Xin Lu, Zhe L. Lin, Hailin Jin, Jianchao Yang, and James Z. Wang, "Rating image aesthetics using deep learning," *IEEE Trans. Multimedia*, vol. 17, no. 11, pp. 2021–2034, 2015.

40 Zhe Dong and Xinmei Tian, "Multi-level photo quality assessment with multi-view features," *Neurocomputing*, vol. 168, pp. 308–319, 2015.

41 Weining Wang, Mingquan Zhao, Li Wang, Jiexiong Huang, Chengjia Cai, and Xiangmin Xu, "A multi-scene deep learning model for image aesthetic evaluation," *Signal Processing: Image Communication*, pp. –, 2016.

42 Shu Kong, Xiaohui Shen, Zhe Lin, Radomir Mech, and Charless Fowlkes, "Photo aesthetics ranking network with attributes and content adaptation," in *European Conference on Computer Vision (ECCV)*, 2016.

43 Yueying Kao, Ran He, and Kaiqi Huang, "Deep aesthetic quality assessment with semantic information," *IEEE Trans. Image Processing*, vol. 26, no. 3, pp. 1482–1495, 2017.

44 Z. Wang, D. Liu, S. Chang, F. Dolcos, D. Beck, and T. S. Huang, "Image Aesthetics Assessment using Deep Chatterjee's Machine," in *International Joint Conference on Neural Networks (IJCNN)*, 2017.

45 Shuang Ma, Jing Liu, and Chang Wen Chen, "A-lamp: Adaptive layout-aware multi-patch deep convolutional neural network for photo aesthetic assessment," in *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, 2017, pp. 722–731.

46 Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States.*, 2012, pp. 1106–1114.

47 K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.

48 Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.

49 Kuo-Yen Lo, Keng-Hao Liu, and Chu-Song Chen, "Assessment of photo aesthetics with efficiency," in *Proceedings of the 21st International Conference on Pattern Recognition, ICPR 2012, Tsukuba, Japan, November 11-15, 2012*, 2012, pp. 2186–2189.

50 Zhe Dong, Xu Shen, Houqiang Li, and Xinmei Tian, "Photo quality assessment with DCNN that understands image well," in *MultiMedia Modeling - 21st International Conference, MMM 2015, Sydney, NSW, Australia, January 5-7, 2015, Proceedings, Part II*, 2015, pp. 524–535.

51 Katharina Schwarz, Patrick Wieschollek, and Hendrik P. A. Lensch, "Will people like your image?," *corr*, vol. abs/1611.05203, 2016.

52 Xinmei Tian, Zhe Dong, Kuiyuan Yang, and Tao Mei, "Query-dependent aesthetic model with deep learning for photo quality assessment," *IEEE Trans. Multimedia*, vol. 17, no. 11, pp. 2035–2048, 2015.

53 Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna, "Rethinking the inception architecture for computer vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, 2016, pp. 2818–2826.

54 Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA.*, 2017, pp. 4278–4284.

55 X. Jin, J. Chi, S. Peng, Y. Tian, C. Ye, and X. Li, "Deep Image Aesthetics Classification using Inception Modules and Fine-tuning Connected Layer," in *The 8th International Conference on Wireless Communications and Signal Processing (WCSP)*, 2016.